**Титульник**

**Слайд 2 - Постановка цели и задач**

Целью данной работы является исследование методов интеллектуального анализа данных в приложении к аналитике фискальных данных с последующим применением этих методов в разработке аналитической системы.

Задачи, которые необходимо выполнить для достижения цели следующие:

* Во-первых - Провести анализ предметной области.
* Во-вторых - Сформулировать задачу анализа фискальных данных с точки зрения бизнеса и с математической точки зрения;
* Также нужно разработать аналитический модуль системы и реализовать возможность интеграции системы с другими системами;

**Слайд 3 - Фискальные данные**

Фискальные данные – информация, содержащаяся в чеках, генерируемых ККТ при выполнении оплаты (возврата) каких-либо товаров или услуг.

Фискальные данные являются потенциальным источником ценной информации, а ОФД, имея доступ к этим данным, кроме функции передачи этих данных в ФНС может производить их анализ, с целью получения ценной информации. Для извлечения ценной информации из большого количества данных, необходимо использовать инструменты интеллектуального анализа данных (Data Mining).

**Слайд 4 - Постановка задачи**

С точки зрения бизнеса задача анализ рыночной корзины

С точки зрения интеллектуального анализа данных, задача анализа рыночной корзины является задачей поиска ассоциативных правил

**Слайд 5 - Алгоритмы поиска частых наборов**

В работе рассмотрены 4 алгоритма.

Из рассмотренных алгоритмов для поставленной задачи наиболее подходящим является Frequent Pattern-Growth, т.к. он логически разделен на два этапа:

* Построение FP-дерева;
* Извлечение частых наборов из FP-дерева.

Это означает, что данные этапы могут выполнятся независимо друг от друга в различные моменты времени.

**Слайд 6 - Use Case диаграмма системы:**

Функциональность, которую необходимо реализовать в системе можно графически представить в виде Use Case диаграммы.

Проектируемая система может взаимодействовать с двумя действующими лицами: пользователем системы и сервисом интеграции с ОФД. Пользователь системы имеет возможность взаимодействовать через API с аналитическим модулем системы:

* Получать представление префиксного дерева;
* Получать частые наборы из текущей версии дерева;
* Получать ассоциативные правила из текущей версии дерева;
* Вручную запускать процесс обновления дерева.

Сервис интеграции с ОФД имеет возможность загружать в систему новые транзакции двумя способами: через API и через сервис очередей.

**Слайд 7 - Архитектура системы**

Источник данных транзакций, обозначенный в схеме как «ОФД», связан с модулем предварительной обработки не на прямую, а посредством очереди сообщений. Данный подход позволит во-первых – оградить модуль предварительной обработки от перегрузки, во-вторых – избежать потери данных, если модуль обработки по какой-либо причине будет недоступен, например, во время обновления системы.

При получении новой транзакции, модуль предварительной обработки будет производить ее обработку следующим образом: извлекать из транзакции требуемые данные, т.е. список элементов (товаров) участвующих в транзакции, преобразовать эти данные к внутреннему формату системы, после чего записывать эту информацию в БД, с сохранением времени записи. На основе сохраненных транзакций будет строиться специальная структура данных, которая позволит эффективно извлекать аналитическую информацию с которой далее будет работать модуль анализа.

**Слайд 8 - База данных системы**

БД предназначена для хранений исходных транзакций и FP-дерева.

Node – дерево.

**Слайд 9 - Проверка алгоритма**

Отладка алгоритма производилась с использованием открытой базы данных транзакций.

Был создан python-скрипт, выполняющий поиск частых наборов с по 1000 транзакций в БД. Для тех же транзакций было выполнено построение дерева и извлечение частых наборов - числа одинаковые.

**Слайд 10 – Заключение:**

В результате выполнения ВКР было проведено исследование методов интеллектуального анализа данных в приложении к аналитике фискальных данных с последующим применением этих методов в разработке аналитической системы.

В ходе разработки был решен ряд задач:

* Проведен анализ предметной области;
* Сформулирована задача анализа фискальных данных;
* Разработан аналитический модуль системы;
* Реализована возможность интеграции системы с другими системами.

Также необходимо отметить тот факт, что результаты проведенного исследования опубликованы в журнале "Вопросы радиоэлектроники".